A Proposal for a National Health Care Identifier

Barry R. Hieb, M.D., Digital Equipment Corporation
representing ASTM E31.12

ABSTRACT

This paper describes a proposed standard for the creation of a national health care identifier being developed under the auspices of the ASTM. The standard includes the desired properties of such an identifier and a description of existing identifier schemes. It includes a proposed identifier scheme, descriptions of how the proposed scheme would function, and an evaluation of how well the proposed scheme meets the properties outlined in the standards document. This paper provides a partial summary of the material contained in the proposed standard.

INTRODUCTION

The establishment of a robust national identifier to meet the needs of health care has been a long-standing goal of the United States medical establishment. Earlier studies have explored the possibilities and limitations of record linkage without the use of a unique identifier (1). The identifier issue has been brought into sharper focus as the implications of the Computer-Based Patient Record are considered (2). Much of the debate on the identifier topic has focused on the merits and deficiencies of using the Social Security Number (3) as the identifier since it is frequently viewed as the only viable candidate for such a function. This article will describe an alternative candidate - the Universal Healthcare IDentifier (UHID.) The UHID is the result of a 2 year standards development process by ASTM committee E31.12 on medical informatics chaired by Dr. Elmer Gabrieli. During the summer of 1994 the standard is undergoing balloting for acceptance as a formal standard (4).

PURPOSE

The purpose of this article is to describe the proposed standard and identify some of its features and benefits. For purposes of accuracy and clarity much of this article will consist of *selective quotes in italics* of portions of the proposed standard. The UHID standard consists of three parts. The initial part defines the properties which a candidate national health care identifier must possess in order to meet the needs of health care in the United States. The second part contains descriptions of some of the existing identifier schemes. The third part defines a recommended implementation of a UHID and provides an evaluation of how adequately it meets the criteria defined in the inital part of the document.

FUNCTIONS

The proposed standard establishes four basic functions which a candidate national health identifier should meet:

This guide sets forth the fundamental considerations for an UHID that can effectively support at least four basic functions:

(1) positive identification of patients when clinical care is rendered;

(2) automated linkage of various computer-based records on the same patient for the creation of lifelong electronic healthcare files;

(3) providing a mechanism to support data security for the protection of privileged clinical information;

(4) enable the use of technology for patient records handling to keep health care operating costs at a minimum.

The proposed standard then goes on to define thirty criteria that describe more detailed requirements for any candidate identifier. Each criterion is given as a titled entry which describes that requirement. Included below are selected entries from the list of criteria contained in the proposed standard.

6-3 Atomic

A UHID should be a single data item. It should not contain subelements which have meaning outside of the context of the entire UHID. Nor should the UHID consist of multiple items which must be taken together to constitute an identifier.
This criterion implies that a collection of data items (e.g. name, date of birth, mother's maiden name) does not qualify as an identifier.

6-5 Content-free

The UHID should not depend on possibly changing or possibly unknown information pertaining to the person.[1]

6-7 Cost-effective

The UHID system chosen should achieve maximum functionality while minimizing the investment required to create and maintain it.

6-9 Disidentifiable

It should be possible to create an arbitrary number of UHIDs which can be used to link medical information concerning specific individuals but which cannot be used to identify the associated individual. These are Encrypted Universal Health Care Identifiers (EUHIDs). EUHIDs should, with the exception of disidentification, have all of the properties attributable to UHIDs including verification (cf. section 6-30, page 12). It should be clear to all users whether a specific identifier represents a UHID or an EUHID. The EUHID scheme should be capable of generating a large number (at least hundreds) of EUHIDs for a single individual.

An EUHID creates an alias which can be used to link information without identifying the corresponding individual[2]. The need for multiple EUHIDs arises from the wide variety of anticipated uses for disidentification and the inability to reconcile conflicting operational requirements between these various uses.

6-10 Focused

The UHID should be created and maintained solely for the purpose of supporting health care. Its form, usage, and policies should not be influenced by the needs or requirements of other activities.

6-12 Identifiable

It must be possible to identify the person associated with a valid UHID. Identifying information

may include such standard items as name, birthdate, sex, address, mother's maiden name, etc. This information is not incorporated in the UHID but is associated with it by linkages.

6-19 Permanent

A UHID, once assigned, remains with that individual. It is never reassigned to another person even after the individual's death.

6-27 Unique

A valid UHID or EUHID identifies one and only one person. A person should have only one UHID. (Note that for purposes of disidentification a person may have an arbitrary number EUHIDs as defined in section 3-11, cf. page 4.)

6-30 Verifiable

A user should be able to determine that a candidate identifier is or is not a valid UHID without requiring additional information. This should support the ability to detect accidental misinformation such as typographical errors. It is not meant to be able to preclude intentional misinformation.

A UHID should have a mechanism - such as check digits - to enable the user to perform a consistency check to ensure that the identifier is valid.

After listing the criteria for an identifier as partially outlined above the standard goes on to discuss several topics that are related to the identifier. The first of these sections discusses the need for Temporary Patient Identifiers to be used when a UHID is not available. This is followed by a discussion on encrypted identifiers (EUHIDs.) EUHIDs are the method whereby the identification scheme supports the disidentification requirement listed in 6-9.

. . . . Essentially an EUHID creates an alias which can be used to link various information items without knowing whose information is being linked. It is generally assumed that such an alias would be used for ordering a single patient care episode, e.g. a sin-

---

[1]Including content in the UHID makes it impossible to assign the "correct" identifier if that information is not known. It also leads to invalid situations if the information changes: e.g., what happens to an identifier based on gender if the person has a sex-change procedure?

[2]Note that the standard spells out the requirements for disidentification but does not attempt to determine the policies and procedures under which such disidentification capability will be used.

gle hospitalization, or a single procedure such as a sensitive laboratory test. As a result, the system must be capable of creating multiple (hundreds or more) EUHIDs to cover potentially large numbers of care episodes for a given individual. . . .

Since EUHIDs are used to provide disidentified patient information linkage it is important that they not contain content relating to the individual. Items such as sex, birthdate, names, etc. must be excluded from EUHIDs to prevent compromising their disidentification function.

An EUHID must be revealable in order to serve its linkage function. Thus, it should be possible to print it on reports, store it in databases, etc. in a manner analogous to a person's UHID without compromising its disidentification function.

[Two examples of possible EUHID use may be helpful. A hospital wishes to order a sensitive test (e.g. HIV) on a patient. The institution obtains an EUHID for the patient and uses it to label the sample. When the results are obtained the institution can then take appropriate action to link the result to the patient depending on the results of the test and the applicable rules and regulations.

In the second example a researcher wishes to obtain a blinded patient population for research. She provides the number of patients needed and the type of information needed on each patient to an authorized health care information agency. The agency gathers the needed information on each patient and links it using an EUHID. The disidentified information is then provided to the researcher for use. If at some later date the information needs to be unblinded this can be done by provinding a link between each EUHID and the corresponding UHID.]

The proposed standard concludes by listing some of the policy decisions which will be required in order to implement a national identification scheme. It is noted that these policy decisions - while essential to the implementation of any national identificaion system - are beyond the scope of the ASTM standards effort.

Two appendices are attached to the proposed standard. The first appendix describes some of the existing identifier systems in use in the United States and Europe. Identifiers reviewed in this section include the Social Security Number, the Swedish "Personal Identity Number", the Danish personal identifier, and the identifier used in Finland. A novel scheme based on geographic information is also included (5).
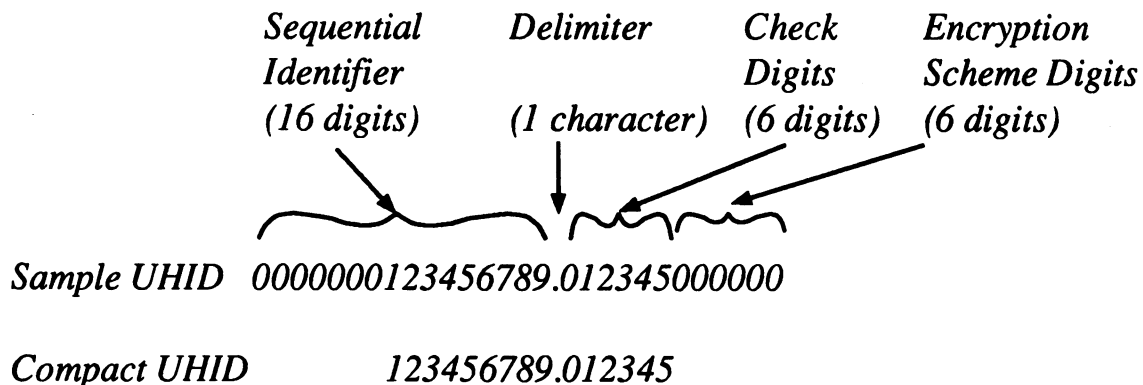
PROPOSED IDENTIFIER SCHEME

The second appendix to the proposed national identifier standard outlines a candidate implementation scheme. It begins by describing the nature and structure of the proposed identifier. Figure 1 provides an overview of the structure of this proposed identifier. It consists of a 16 digit Sequential Identifier (SI), a single character delimiter, 6 check digits, and 6 encryption digits. The SI provides a unique number for each individual. The delimiter marks the boundary between the SI and the check digits. The check digits implement an error detection scheme which is able to ensure the validity of the UHID to a certainty level of one part in one million. The encryption digits provide the ability to create up to one million EUHIDs for each person. As shown in Figure 1, a full identifier constitutes 29 digits but leading and trailing zeroes may be truncated to provide a compact identifier which would typically be on the order of 16 or 17 digits.

Appendix 2 of the proposed standard goes on to briefly describe each of the subcomponents of the UHID and provide usage examples showing how the proposed identifier would be used to support three typical activities: assigning a UHID to a person, generating an EUHID, and decrypting an EUHID,

The proposed UHID uses check digits to support the verification requirement stated above in section 6-30. The next section of the appendix discusses how these check digits are computed for both UHIDs and EUHIDs. Tables needed for this computation are included. The final section of appendix 2 contains a preliminary evaluation (by one of the authors) of the candidate identifier against the 30 criteria listed in section 1 of the document. Each criterion is evaluated on the following scale:

1 - not supported
2 - minimally supported
3 - inadequately supported
4 - adequately supported
5 - fully supported
X - cannot be rated

471

Fig. 1 - Sample UHID Format

Sequential Identifier (16 digits)	Delimiter (1 character)	Check Digits (6 digits)	Encryption Scheme Digits (6 digits)

Sample UHID   0000000123456789.012345000000

Compact UHID   123456789.012345

The summary of this evaluation is provided in the following table.

UHID EVALUATION SUMMARY

Evaluation Category	Number of Criteria
1	0
2	0
3	2
4	5
5	18
X	5

From this evaluation it can be seen that the proposed scheme appears to adequately meet all but two of the criteria listed in the standard. The two criteria which were deemed to be inadequately supported were cost (because creation of a new UHID system would likely require a significant financial investment) and the limited ability to "split" an identifier which has inadvertently been assigned to two individuals.

CONCLUSION

The debate concerning the future of health care in the United States makes the present time particularly appropriate for an evaluation of the need for a national health care identifier. Most discussion on this topic to date has been focused on the merits and demerits of the use of the Social Security Number, largely because of the lack of any viable alternative.

Many disadvantages of the SSN have been identified. However, it has the perceived advantages of cost-effectiveness and rapid implementation. The UHID scheme discussed in this proposed standard provides the opportunity to open a debate on the question of whether it is feasible to create a more functional identifier in a timely and cost-effective manner.

REFERENCES

1. Newcombe, Howard B.: Handbook of Record Linkage; Oxford University Press, Oxford, England, 1988.

2. Health Data in the Information Age Use, Disclosure, & Privacy. Molla S. Donaldson & Kathleen N. Lohr, editors, Committee on Regional Health Data Networks, Division of Health Care Services, Institute of Medicine, Washington, D.C., National Academy Press, 1994.

3. Federal Register, November, 1922, The Social Security Number, policy and general procedures.

4. ASTM: Guide for the Properties of A Universal Healthcare Identifier. (Proposed standard in balloting process. To obtain a copy contact Dr. Elmer Gabrieli, CBML, 4 Cambridge Center, Cambridge, Mass. 02139)

5. Carpenter, Paul and Chute, Christopher: The Universal Patient identifier: A Discussion and Proposal, Proceeding of the 17th Annual Symposium on Computer Applications in Medical Care, p. 49-53, 1994.